

## \* NOTICES \*

JPO and INPIT are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.\*\*\*\* shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

---

## DETAILED DESCRIPTION

---

[Detailed Description of the Invention]

[0001]

[Field of the Invention]This invention relates to the face image registration device which extracts a face area out of the inputted picture, and carries out dictionary registration of the face automatically, or its method.

[0002]

[Description of the Prior Art]Face image recognition art is a human interface or a component engineering indispensable to construction of a security system. Face image recognition consists of face area extraction, face-characteristic-points extraction, and discernment. About the latest technology trends, literature [latest research trend [ of Mitsuru Shiono and Hidehiko Sanada" personal authentication art ]", Shingaku Giho OSF92-17.] is detailed. The conventional face picture identifying method can be classified into two of the roughly divided followings.

[0003]One is the method of calculating similarity with a target person's feature vector which parameterizes the position of the focus, such as eyes, a nose, and a mouth, shape, and size, generates a feature vector, and is registered beforehand. The person showing a dictionary vector with the highest similarity is discriminated from a person in question. These are classified into the structural analysis technique.

[0004]Another method is a method based on the similarity of the pattern of the picture which it was located [ picture ] by geometrical conversion of the two-dimensional affine conversion on the basis of the focus, such as a pupil and a nose, etc., and had size normalized, and the dictionary normalized image registered beforehand. The person who expresses a dictionary image with the highest similarity like the former is discriminated from a person in question. These can be classified into the technique like a pattern according to the method put in practical use in the conventional character recognition.

[0005]In both [ of the two above-mentioned techniques ] cases, statistical distance, such as a correlation value between pictures and Euclidean distance in the inside of a feature space, is used as similarity. Methods, such as various pattern recognition theories put in practical use in character recognition, for example, a subspace method, [ Hidemitsu Ogawa written by Erkki Oja and Sato \*\*\*\* and "pattern recognition and subspace-method" Sangyo Tosho Publishing (1986)], and compound similarity, are applicable to this identification processing. The composition of the identifying method is detailed to [Funakubo \*\* "pattern recognition" KYORITSU SHUPPAN (1991)], [Taizo Iijima "pattern recognition theoretical" Morikita Shuppan (1989)], etc.

[0006]In addition to the above-mentioned facial recognition processing, on the occasion of actual application of the above-mentioned facial recognition, a user's registration, i.e., a registrant's face dictionary generation, becomes indispensable. For example, in the handwriting recognition of Chinese character using a subspace method, the dictionary was generated from the study sample picture of hundreds of sheets for every character. A three-dimensional object like the target face by this invention, Compared with a character, change of shape, such as lighting conditions, for faces, and an expression change, or luminosity is large, Still a lot of study sample pictures are needed [Hiroshi Murase, Shelley Nay Ya, "the three-dimensional object recognition by two-dimensional collation", IEICE TRANSACTIONS (D-II) J77-D-II, 11, and pp. 2179-2187-1994.].

[0007]The conventional dictionary generation Literature [Shigeru Akamatsu, Tsutomu Sasaki, Akio Fukamachi, Yasuhito Suenaga, "the robust method of identifying a transverse-plane face by shade image matching", IEICE TRANSACTIONS (D-II), J76-DII, 7, pp.1363-1373, 1993.], [Yoshie Komatsu, Yasuo Ariki, "logging recognition of the face by the direction using a subspace method", PRU95-191, pp. 7-14, and 1996.], [M. Turk, A. P. Pentland : " Facerecognition using eigenfaces", Proc. CVPR 11, pp. 453-458, and 1993.], [Alex Pentland,Baback Moghaddam, Thad Starner, " View-based and modular eigenspaces for face recognition " ,CVPR'94, The help was performing logging of a face and selection of the picture from the picture which turned and photoed the registrant's face under good lighting conditions in predetermined, for example, a transverse plane, 15 degrees of right and left, and the direction of 15 degrees of up-and-down \*\* so that pp.84-91, 1994.], etc. might see.

[0008]A face tends to produce a physiological change of the influence of the condition of the day, an expression change, hair, and a mustache, etc. Therefore, in order to realize a high recognition rate, it is necessary to update a face dictionary for every fixed time. However, the frequent renewal of a dictionary will raise a user's burden. So, in order to make facial recognition easy to use, it becomes important how a burden is eased and registered without making a user conscious.

[0009]For example, when it is interrupted temporarily and leaves the work in an information

terminal, the function to prevent others' access is considered. although using the dictionary generated beforehand is also considered -- the person himself/herself -- it is more effective to use the dictionary image generated as much as possible immediately before, in order to raise identification and an others exclusion rate However, in order to make a user's burden high, dictionary registration of carrying out dictionary registration to the degree of a leaving chair is carried out without a user's making it conscious while accessing information machines and equipment.

[0010]However, the study sample collected without making a registrant conscious contains the data which is not effective as learned data the case where eyes are usually closed, when expression is changing a lot. If dictionary generation is performed from the study sample in which these error pictures are included, the accuracy of a dictionary will fall. In order to raise dictionary accuracy, it is necessary to increase the number of learned data, and real-time dictionary generation cannot be coped with in selection by a help.

[0011]

[Problem(s) to be Solved by the Invention]Conventionally, the help was performing generation of the face dictionary using predetermined, for example, a transverse plane, 15 degrees of right and left, and the face picture that was made to turn a registrant's face in the direction for 15 degrees of every upper and lower sides, and was photoed a priori under good lighting conditions. A face changes with time. Therefore, in order to realize a high recognition rate, it is necessary to update a face dictionary for every fixed time. However, performing frequent renewal of a dictionary will raise a user's burden.

[0012>About the automatic study sample collection for dictionary generation, if the method [Japanese Patent Application No. No. 61463 [ eight to ]] of having applied already is applied, it is realizable. However, the technical problem that much data which is not effective in dictionary generation is contained remains in the learned data collected without still making a registrant conscious. In order to raise dictionary accuracy, it is necessary to increase the number of learned data, and real-time dictionary generation is difficult to realize in selection by a help.

[0013]Then, this invention provides a user with a burden, \*\*\*\*, a face image registration device that can eliminate registration of an unnecessary face picture and can raise dictionary accuracy, and a method for the same.

[0014]

[Means for Solving the Problem]A face image registration device whose this invention is characterized by that a face image registration device comprises the following and which has a dictionary means by which a person's face picture is registered.

An image input means which inputs said person's picture.

A face area extraction means to extract face image input which is a picture of said person's face area out of a picture inputted by said image input means.

A decision means which judges whether face image input extracted by said face area extraction means is a face picture which can be registered into said dictionary means.

A registration means made to register into said dictionary means by said decision means by making into said face picture face image input judged that registration is possible.

[0015]A face image registration method for registering a person's face picture whose this invention is characterized by that a face image registration method comprises the following.

An image input step which inputs said person's picture.

A face area extraction step which extracts face image input which is a picture of said person's face area out of a picture inputted in said image input step.

A judgment step face image input extracted in said face area extraction step judges it to be whether it is a face picture in which said registration is possible.

A recording step into which face image input judged that registration is possible in said judgment step is made to register as said face picture.

[0016]It is judged whether it is effective as learned data by comparing a standard beforehand set up in characteristic quantity which can be found from a face area extracted, for example as it is the above-mentioned invention. As characteristic quantity, a picture which normalized on the basis of a size of face characteristic points, such as a pupil, a nostril, and a mouth edge, quantity which can be found from position information, or these points is used. It becomes possible to remove automatically a picture from which expression changed a lot by this, a picture which closed eyes, a picture which opened a mouth, etc. from a learning image.

[0017]Furthermore operation of predetermined [, such as mouse operation, keyboard grabbing, and button grabbing, ] is detected, and dictionary generation is performed only for a picture inputted during operation. Time of useless learned data collection decreases for being aimed only at in operation, and it is efficient. A possibility of collecting pictures which are not effective as dictionaries in a case of having turned to width, etc. also decreases. According to this invention, a device and a method of extracting a face area out of an inputted picture, and carrying out dictionary registration automatically are realizable.

[0018]

[Embodiment of the Invention]Hereafter, one example of this invention is described based on drawing 7 from drawing 1.

[0019]In information terminal equipment, such as a personal computer, when interrupting under work temporarily and leaving a terminal, it is a required function to prevent others' access. Therefore, this example explains taking the case of the information-machines-and-equipment terminal unit 10 which carries the function to prevent others' access at the time of a leaving chair.

[0020]Drawing 1 is a block diagram showing the outline of the information terminal equipment 10 with security concerning this example.

[0021]This device 10 comprises the image input part 11, the face region extracting part 12, the situation-recognition part 13, the private seal 114, the dictionary generation part 15, the access control section 16, the dictionary image 17, and the prescribed operation primary detecting element 18.

[0022](Image input part 11) The image input part 11 is for inputting the picture of the person who becomes a recognition object, for example, consists of TV cameras. The picture 01 inputted from this image input part 11 is digitized by the A/D converter, and is sent to the face region extracting part 12. For example, a TV camera is installed in the lower part of a monitor of the information terminal equipment 10. Or it may install in the rectangular head of a monitor.

[0023](Face region extracting part 12) The face region extracting part 12 always continues extracting the face area picture 02 from the inputted image sent from the image input part 11. Let the field which calculates a correlation value and has the highest correlation value be a face area in this example, moving the standard face image (template) registered beforehand over the full screen. Let the partial maximum points of a correlation value be a face area candidate. Suppose that a face does not exist, when the correlation value in this face area candidate is lower than the set-up threshold. A maximum correlation value is specifically set to 100 as a threshold, and it is set as 30. If two or more templates are used by compound similarity etc. in order to correspond to direction change of a face, a face area can be extracted still more stably. This processing may be transposed to the extraction method based on the color information described previously.

[0024](Situation-recognition part 13) In the situation-recognition part 13, a user's condition (under work, a leaving chair, taking a seat) is discriminated from a time change of a face area extraction result, and the change of dictionary generation mode and recognition mode is performed.

[0025]It is set as recognition mode in the periods from T0 to T1 the time of a face area beginning to be detected as shown in drawing 2. the recognition processing described in this mode later -- the person himself/herself -- identification processing is performed. It becomes accessible when identified the person himself/herself at the T1 time.

[0026]When it changes to dictionary generation mode after T1 time and the dictionary generation signals from the prescribed operation primary detecting element 18 are inputted, collection of learned data is performed. If it becomes impossible to detect a fixed time face area, it will judge that the leaving chair was carried out and a screen lock will be started immediately. Or a screen lock may be started, after conveying the main point to a user with light, a sound, etc. and seeing a user's reaction. The time T1 can be changed according to the kind of an object model or access operation here.

[0027](Recognition part 14) The recognition part 14 consists of the face-characteristic-points extraction part 14a, the normalized image generation part 14b, and the pattern matching part 14c. The block diagram of the recognition part 14 is shown in drawing 3.

[0028]In the face-characteristic-points extraction part 14a, the focus, such as a pupil, a nose, and \*\*\*\*, is extracted from the inside of the extracted face area. The method [Japanese Patent Application No. No. 61463 [ eight to ]] which combined the configuration information for which it has already applied, and pattern information is applicable.

[0029]The fundamental idea of this method asks for the candidate of the focus by configuration information with high accuracy of position, and verifies it by pattern matching. Since this method positions by configuration information, it can expect high accuracy of position. Since matching which used the multi-template for selection of the right focus from a candidate group is applied, it is robust to change of the shape luminosity of the focus. About processing speed, since pattern matching is carried out only to the candidate who narrowed down with the degree-in-separation filter with little calculation cost, compared with the method of carrying out the pattern matching of the whole, drastic reduction of computational complexity is realizable.

[0030]In addition to this, To edge information. Based method [A. L.Yuille, " Feature extraction from faces using deformable templates " , IJCV, vol. 8: 2, pp. 99-111, 1992.] [Shizuo Sakamoto, Miyao -- Yoko -- Tajima -- Joji -- " -- a face -- a picture -- from -- an eye -- the focus -- extraction -- " -- IEICE TRANSACTIONS -- D-II -- Vol . -- J -- 76 - D-II -- No . -- eight -- pp . -- 1796 - 1804 -- August -- 1993 . --] -- intrinsic space -- a method -- having applied -- Eigen. feature method [Alex Pentland, Baback Moghaddam, Thad Starner, " View-based and modular eigenspaces for face recognition ", CVPR'. The method [Tutomu Sasaki, Shigeru Akamatsu, Yasuhito Suenaga, "the method of alignment of the face using the sexual desire news for face image recognition", IE91-2, pp. 9-15, and 1991.] of being based on 94, pp. 84-91, 1994.], and color information is applicable.

[0031]In the normalized image generation part 14b, it normalizes on the basis of the focus. The example of the normalizing process on the basis of a pupil and a nostril is shown in drawing 4. Direction of vector E1E2 is amended in parallel, and, as for 2 times vector E1E2 and a dip, the center coordinates of a normalized image and breadth set up point CP of 1/3 by 2 twice vector c1c2 from on c1 and c2 further.

[0032]In the pattern matching part 14c, it asks for pattern similarity as compared with a normalized image and the face picture currently stored in the dictionary image. When pattern similarity is higher than a reference value, it identifies being the person himself/herself. In being smaller than a standard, it considers it as others. As the method of pattern matching, a correlation technique, a subspace method, a compound similarity method, etc. are applicable.

[0033](Prescribed operation primary detecting element 18) The prescribed operation primary detecting element 18 detects whether the user is performing operation predetermined [, such

as a keyboard, a button push, or operation of a mouse, ]. The dictionary generation part 15 is controlled based on a detection result.

[0034] Since a possibility that the transverse-plane picture of a user's face is reflected to the image input means 11 is high when it specifically detects that the user is performing operation predetermined [, such as keyboard grabbing, a button push, and operation of a mouse, ], Since it is suitable to carry out dictionary registration, while performing predetermined operation, dictionary generation signals are transmitted to the dictionary generation part 16.

[0035] (Dictionary generation part 16) The dictionary generation part 16 consists of the face-characteristic-points extraction part 14a, the frame evaluating part 16a, the normalized image generation part 14b, the learning image storage parts store 16b, and the principal-component-analysis part 16c. With the recognition part 14, the face-characteristic-points extraction part 14a and the normalized image generation part 14b are common, and are used here. The block diagram of the dictionary generation part 16 is shown in drawing 5.

[0036] As for dictionary generation, if dictionary generation signals are inputted from the prescribed operation primary detecting element 18, collection of learned data will be performed.

[0037] In the frame evaluating part 16b, it is judged whether the face area picture 02 extracted from the information on the extracted face characteristic points is effective in dictionary generation. Although three valuation bases are applied here, it may apply in individual or combination.

[0038] (i) Specify the physical relationship of each focus beforehand, and when this physical relationship collapses, judge that it is not effective as learned data. For example, it is the physical relationship of both pupils, a nostril, and a mouth edge, etc.

[0039] (ii) Make evaluation the inherent face picture and pattern similarity which were stored in the dictionary image 17. When similarity is smaller than a standard, it is judged that it is not effective. The homogeneity of learned data will be improved, so that a standard is made high. When too high, it becomes impossible however, to correspond to the change for faces etc.

[0040] Or the dictionary image generated from the learning image normalized in the right focus a priori instead of the inherent face picture stored in the dictionary image 17 may be used.

[0041] If it furthermore takes into consideration to for faces, it is possible to remove the face etc. which turned to width extremely. In this case, when comparison with the face dictionary pattern currently beforehand prepared for direction 1 - for [ N / every ] faces is performed and it is in agreement with the predetermined dictionary for faces, it is judged that it is effective in dictionary generation.

[0042] (iii) Detect the learned data which investigated the relation of the focus extracted for every frame, and closed eyes. The coordinates of the focus extracted newly can be expressed by the linear combination of the coordinates of the focus already extracted with the frames 1-4

before that. Therefore, if the focus is correctly extracted with all the frames, the error at the time of carrying out linear combination will become small. Conversely, an error becomes large, when eyes are closed and eyebrows and a pupil are mistaken.

[0043] This is explained using drawing 6 in detail.

[0044] If an orthogonal projection model is assumed, the two-dimensional coordinates of the face characteristic points seen from arbitrary directions, without having a three-dimensional model, By the linear combination of the two-dimensional coordinates of the focus matched on the picture of four sheets as shown in the formula (1) and (2). [S. Ullman, R. Basri:

"Recognition by Linear Combinations of Models, IEEE Trans. PAMI, Vol. 13, No. 10, pp. 992-1006 which can be expressed, 1991.] [Yasuhiro Mukogawa, Yuichi Nakamura, Tomoichi Ota, "generation of the face picture of the arbitrary direction using two mug shots", an information processing theory, Vol. 37, No. 4, pp. 635-644, and 1996.]. The two-dimensional coordinate value (X, Y) of the focus newly detected from two-dimensional coordinate value [ of the right focus in four frames already extracted in the video sequence using this character ]  $(x_1, y_1)$  -  $(x_4, y_4)$  is verified.

[0045] Linear combination coefficient  $a_i$ ,  $b_i$ , and  $(i=1, 4)$  are calculated with a least square method from the two-dimensional coordinates of the focus to the picture of four frames, and new coordinates (X, Y). The error Res which calculates  $(x', y')$  for approximate coordinates from the coupling coefficient obtained conversely (X, Y) is calculated shortly.

[0046]

[Equation 1]

$$X = a_1 \cdot x_1 + a_2 \cdot x_2 + a_3 \cdot x_3 + a_4 \cdot x_4 \quad (1)$$

$$Y = b_1 \cdot y_1 + b_2 \cdot y_2 + b_3 \cdot y_3 + b_4 \cdot y_4 \quad (2)$$

$$Res = \sqrt{(X - x')^2 + (Y - y')^2} \quad (3)$$

The error Res shows the coordinates error of the new focus. Since the error has arisen in the extraction position when an error is larger than a threshold, this data is not effective as learned data.

[0047] It is also possible to judge whether distribution of the average luminance of a face area or luminosity is contained in the reference interval besides the valuation basis of the above-mentioned (i) - (iii).

[0048] The learned data judged that the learning image storage parts store 16b is the above-mentioned validity is stored. If the stored number of learned data reaches the number of regular, a learning image will be sent to the principal-component-analysis part 16c.

[0049] In the principal-component-analysis part 16c, principal-component-analysis (KL development) application is carried out to the image data stored in the learning image storage

parts store 16b, and it asks for eigenimage. Top N eigenimages are stored in the dictionary image 17 from the one where a characteristic value is larger.

[0050]Although the above-mentioned example explained an example which performs dictionary generation in real time from video, it is also possible to carry out, once it stores in a memory only a frame by which a detecting signal from a prescribed operation primary detecting element came from video. In this case, since possibility effective in dictionary generation is aimed only at a high frame, improvement in the speed and stability of processing can be measured.

[0051]Although a subspace method which is a technique like a pattern is applied to discernment in the above-mentioned example, it may transpose to a structural technique.

[0052](Explanation of operation of this device 10), next operation of this example are explained along with drawing 7.

[0053]The first access performs the usual login which enters a password etc. Or it is also possible to perform automatic login using a dictionary of a user who registered a priori.

[0054]In the face region extracting part 12, a face area is extracted from a picture first inputted from the image input part 11, and it is sent to the situation-recognition part 13. In the prescribed operation primary detecting element 18, prescribed operation detecting signals, such as mouse operation, keyboard grabbing, and button grabbing, are sent to a situation-recognition part. In the situation-recognition part 13, a change in recognition mode and dictionary generation mode is performed based on time existence of a face area. It is controlled whether learned data is collected based on a prescribed operation detecting signal.

[0055]T1 is first set as recognition mode from the time T0 -- the last user -- it is identified whether you are the person himself/herself.

[0056](i) A recognition mode face area is sent to the face-characteristic-points extraction part 14a, and a pupil, a nostril, and a mouth edge are detected. Next, normalization is performed on the basis of these focus. Similarity with a user's dictionary image is calculated last time by the pattern matching part 14c, and when similarity is higher than a reference value, a normalized image is identified the person himself/herself who was using it last time, and becomes accessible. Access is forbidden when smaller than a reference value.

[0057]Next, from the time T1, it is set as dictionary generation mode.

[0058](ii) When the dictionary generation mode prescribed operation primary detecting element 18 detects prescribed operation and a prescribed operation detecting signal is set to ON, collect learned data and generate a dictionary image. A face area picture extracted from a picture inputted continuously is sent to the face-characteristic-points extraction part 14a, and a pupil, a nostril, and a mouth edge are detected. Extracted feature point information is sent to the frame evaluating part 16a, for example, an error is calculated from a formula (3). When an error is smaller than a threshold, it is judged that it is effective as learned data, and it is stored

in the learning image storage parts store 16b. If learned data more than regulation number of sheets is stored in the learning image storage parts store 16b, learned data will be sent to the principal-component-analysis part 16c, and a dictionary image will be generated.

[0059]If a user does a leaving chair, a screen lock will be started immediately. If a user takes a seat again, it will return to the first processing.

[0060]In the above-mentioned explanation, when the prescribed operation primary detecting element 18 detected prescribed operation and a prescribed operation detecting signal was set to ON, a dictionary image was generated, but further, after a prescribed operation detecting signal suits at OFF, a dictionary image may be continuously generated between fixed time (for example, for 1 minute).

[0061]An example in service of not only security of (an example of change), next an information terminal but an information terminal is described.

[0062]For example, when registering a face, it considers memorizing simultaneously not only characteristic quantity obtained from a face area but an operation history of an information terminal. while operating discernment of the human being, simultaneously a terminal information on a face, simultaneously by registering, it can extract from a history what kind of information was searched for.

[0063]If information retrieval is made into an example, a user will do work which inputs a search formula etc. In addition to information required for identification, a history of the search formula and the contents of search is stored in a database. Based on stored data, the user's interesting contents are presumed and extracted and it holds by performing new information retrieval.

[0064]And when a user performs login to an information terminal, and an operation return again, service which provides automatically information considered to be the request newly searched to a terminal is attained.

[0065]By linking such hysteresis information with registration information on a face, and being stored, it may become instead of a time stamp when to have retrieved such information, it may recognize whether it is close to a state of a face at the time of when based on this, and information based on the hysteresis information of those days may be shown.

[0066]As a verification mechanism in the case of registration of a face, if expression detection and recognition are used, it is also possible to take an operation history of information according to human being's emotion state, and a dictionary at the time of usual, a dictionary when it laughs, a dictionary when angry, etc. can also be generated. When newly performing facial recognition, an offer of information corresponding to expression information may be performed.

[0067]When these perform continuously renewal of face registration, and renewal of a manipulation-of-information history simultaneously, a clearer offer of information becomes

possible.

[0068]Although these examples explained an information terminal to an example, they may be transposed, for example to home electronics, such as TV and a telephone.

[0069]In the case of TV, a subminiature camera is carried at a remote control. Predetermined operation is defined as operation which is pushing a button of the remote control. Operation which in the case of a telephone micro CCD carries in a receiver and raises a receiver is defined as predetermined operation.

[0070]The device 10 of this example may be used for an automatic teller's machine (henceforth ATM).

[0071]That is, by using this device 10 for a user's personal identification, it forgets based on others and an unauthorized use of usurpation of cash, etc. can be prevented certainly. In this case, the image input part 11 is arranged near the navigational panel of ATM, and it is made for a user's face to be reflected certainly. An account number and renewal registration time may be registered simultaneously with a face picture.

[0072]

[Effect of the Invention]As mentioned above, in order to register only the face image input which extracted the new face picture out of the inputted picture, judged whether it was a face picture to which this deserves dictionary registration, and was judged to be effective according to this invention, registration of an unnecessary face picture can be eliminated and dictionary accuracy can be raised.

---

[Translation done.]